

NASA's Earth Science Data Systems Standards Endorsement Process

Richard E. Ullman
NASA/ Goddard Space Flight Center
Greenbelt, MD USA
richard.e.ullman@nasa.gov

Yonsook Enloe
SGT Inc.
Greenbelt, MD USA
yonsook@mindspring.com

Abstract—Starting in January 2004, NASA instituted a set of internal working groups to develop ongoing recommendations for the continuing broad evolution of Earth Science Data Systems development and management within NASA. One of these Data Systems Working Groups is called the Standards Process Group (SPG). This group's goal is to facilitate broader use of standards that have proven implementation and operational benefit to NASA Earth science. The process is fundamentally different from past standards strategies in several ways. The focus on communities of practice to propose and comment is a key component. So too, is the emphasis on endorsement of demonstrated practices as an additional source of common practices apart from the traditional standards making bodies. Only after practices have been judged to have useful implementation and beneficial operation will they be endorsed.

The Standards Process Group has already conducted a community review of the DAP 2.0 as a data transport mechanism and has served as a facilitator for the initiation of science content discussions in several fields. This paper will describe NASA's Earth science data systems standards process and the experience to date.

Keyword-scomponent; best practices; standards

I. BACKGROUND

NASA's Earth science programs have made a substantial investment in the development of data and information systems. This is most evident in the Earth Observing System (EOS) Data and Information System (EOSDIS) Core System (ECS), a multi-mission, geographically distributed, active archive system that presently holds over two petabytes of Earth observation data products. ECS is the central and largest of NASA's Earth science data systems, but NASA also has specialized data systems such as heritage data systems, Earth Science Information Partnership (ESIP) data systems, pathfinder data systems measurement focused data systems and others. Lessons learned in the development of the ECS and the ESIP experiment indicate that data systems and services operate with most efficiency when they are focused in scope and under the management of a domain specialist, that is, a science team or application provider. In future, NASA's Earth science data systems will be even more heterogeneous and distributed. The data systems will be under domain specific management with local siting, implementation and operational decisions. Each system will necessarily rely on the services of other Earth science systems, and often several others, to meet its obligations. This heterogeneity coupled with mutual

dependence can only be possible with adherence to standards. Data and interface standards are one aspect.

II. HOW WE DESIGNED THE STANDARDS PROCESS

Pre-Standards Process Group (SPG) formulation studies [2] examined best practices in the field of developing and choosing standards and lessons learned by NASA projects and others. For the standards studies, we examined lessons learned in the use of standards by heritage missions and the processes, experiences and "cultural factors" of standards making bodies.

The SPG considers endorsement of data systems practices. These are practices, technologies, or standards that increase the ability of NASA-generated data to be shared within and among communities of interest. Such practices include software application interfaces, data and metadata model conventions, data and information identification, common data services, formats, and other related technologies. The SPG is not concerned with practice or application of science or hard engineering *per se*. Proper scientific method, calibration and modeling and the correct design and characterization of observation systems are essential to the usability of resulting data, but are outside the scope of the SPG.

We use the term "communities of interest" or simply "communities" to include various stakeholders and affected constituents. Example communities include science discipline groups, users of similar applications, data systems developers, ESE mission planners, Earth science educators, and others. Membership in each "community" often overlaps the others, but in our use, community is a group of individuals that have both confluent needs and collegial identification.

A critical finding is that standards are best when they grow out of specific needs that are collectively identified by affected communities. These needs become most apparent in the course of particular application. For example, when scientists need to share data with each other, they may find that particular organization of the data, descriptive metadata or interface method is especially valuable. When these scientists can form consensus on those aspects that need standards and what those standards are, then that standard becomes valuable in their particular science community. Such standards naturally evolve in every community of users and when they do, adherence to the standard becomes a natural and expected contribution to the community. A disparity arises because there is not a single community. Different user communities of the same data have different needs and different standards expectation. Often, the

| Report Documentation Page | | | | Form Approved OMB No. 0704-0188 | |
|--|------------------------------------|-------------------------------------|--|---|------------------------------------|
| Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. | | | | | |
| 1. REPORT DATE 25 JUL 2005 | | 2. REPORT TYPE N/A | | 3. DATES COVERED - | |
| 4. TITLE AND SUBTITLE NASA's Earth Science Data Systems Standards Endorsement Process | | | | 5a. CONTRACT NUMBER | |
| | | | | 5b. GRANT NUMBER | |
| | | | | 5c. PROGRAM ELEMENT NUMBER | |
| 6. AUTHOR(S) | | | | 5d. PROJECT NUMBER | |
| | | | | 5e. TASK NUMBER | |
| | | | | 5f. WORK UNIT NUMBER | |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) NASA/ Goddard Space Flight Center Greenbelt, MD USA | | | | 8. PERFORMING ORGANIZATION REPORT NUMBER | |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) | | | | 10. SPONSOR/MONITOR'S ACRONYM(S) | |
| | | | | 11. SPONSOR/MONITOR'S REPORT NUMBER(S) | |
| 12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release, distribution unlimited | | | | | |
| 13. SUPPLEMENTARY NOTES See also ADM001850, 2005 IEEE International Geoscience and Remote Sensing Symposium Proceedings (25th) (IGARSS 2005) Held in Seoul, Korea on 25-29 July 2005. , The original document contains color images. | | | | | |
| 14. ABSTRACT | | | | | |
| 15. SUBJECT TERMS | | | | | |
| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT UU | 18. NUMBER OF PAGES 4 | 19a. NAME OF RESPONSIBLE PERSON |
| a. REPORT unclassified | b. ABSTRACT unclassified | c. THIS PAGE unclassified | | | |

standard expectations of one community clash with the expectations of another.

The task for the SPG in directing evolution of standards is two-fold. First is to accelerate the natural development of standards, and second is to mitigate the disparity of standards from one community of users to another.

III. THE SPG STANDARDS PROCESS

NASA's Earth science data systems standards process must facilitate interoperability between components of the Earth science network of data systems. Establishment of appropriate standards enables flexibility as future data and service providers have well-defined access points to join the network of data systems. This flexibility is central to supporting the evolving strategies of the NASA's Earth science program. To accomplish these goals, NASA's Earth science data systems Standards Process focuses on endorsement of practices that are relevant to Earth science network of data systems and that have mature implementations and proven operational benefit. See the website [3].

A. Organization

The Standards Process Group (SPG): This is the decision-making board of the standards process composed of part time permanent members from NASA's program office, Earth science mission projects, Earth science funded data systems awardees and representative from other agencies. The SPG meet regularly to decide issues and delegate tasks. The SPG is responsible to consider public comments and technical factors in making decisions.

Technical Working Groups (TWG): These are groups commissioned by the SPG to conduct public review and evaluation of candidate standards, related implementations and operational experience. Membership in a TWG is partially drawn from the Standards Process Group membership and partly drawn from technical area experts and/or community members. The duration of a TWG corresponds to the review schedule set by the SPG for a particular candidate standard or Technical Note.

B. Path to RFC

The term "RFC" stands for "Request for Comment". The content of an RFC is either a technical note or a proposed standard. A technical note is any information that the submitter considers significant to the use of a practice within NASA's Earth science programs. For the remainder of this discussion, an RFC can be considered to be a proposed standard.

RFCs can come from any source including individuals that may be associated with or represent NASA's Earth science funded activities, industry, users of Earth science data. The requirements for an RFC will be the same in each case. The proposed standard must be described in sufficient technical detail (or else references that describe the standard must be given); the domain and use of the standard must be clearly defined; citations of successful implementation and operational experience must be provided. In some cases, the SPG may solicit an RFC. Other times, members of the community will

bring forward an RFC to formalize NASA recognition or broaden use of standards that are used in their community.

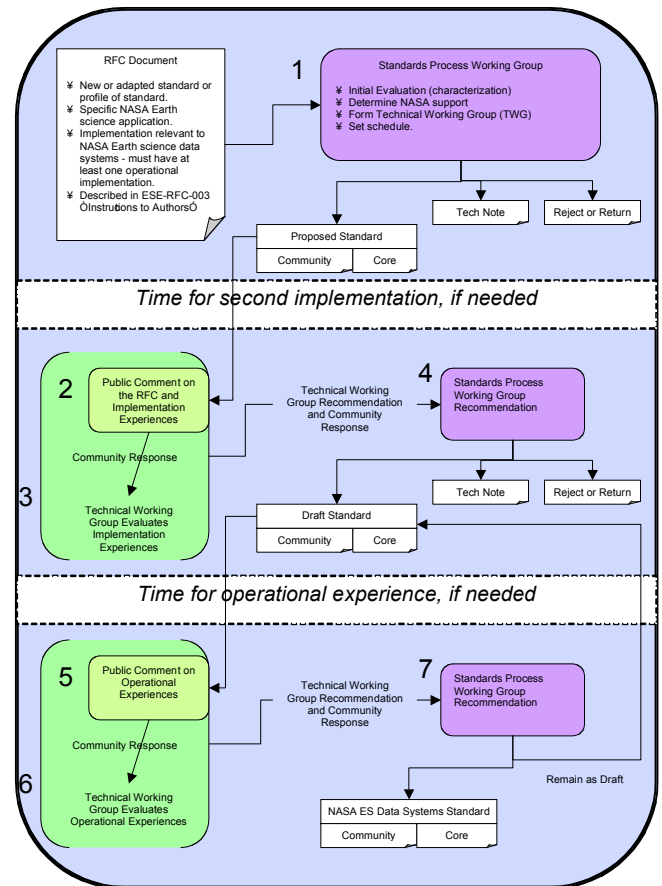


Figure 1. Standards Endorsement Process

C. Path to Endorsement

Figure [1] shows the steps from an RFC to endorsement as an NASA Standard. The process is characterized by technical analysis, open public review and demonstration that the proposal "works". The first step is for the SPG to perform an initial screening and characterization. A TWG is assigned and a schedule is set, taking into consideration NASA need dates and support commitments. Also, any RFC must have two or more implementations before it can advance to draft status.

The TWG and the public review the "Proposed Standard" focusing on the significance of the implementations. In this context, implementations are independent and interoperable uses of the practice. For example, if the RFC proposes a format for a particular class of science product, demonstration of the use of that format by two separate "implementing organizations" would be considered two implementations. The SPG considers the response and decides if the RFC should advance to "Draft Standard"

The review and decision process is repeated, but this time the focus is on operational experience. Not only must the standard be demonstrated to work, but also the standard must be shown to work under conditions that are judged by the TWG

and ultimately the SPG to be significant operationally before a RFC can advance to “NASA Earth Science Standard” status.

The process defines a number terms for status of standards in the process, they are related to the strength of endorsement.

- **RFC:** Any proposal to endorse a practice for NASA Earth science data systems use. These may be standards or technical notes.
- **Proposed Standard:** The SPG presents an RFC for public community review and comment.
- **Draft Standard:** The proposal has technical merit, has been implemented more than once and may become endorsed as a NASA Earth science standard.
- **NASA ES Standard:** There is significant operational experience to endorse this as a NASA Earth science standard. NASA funded data systems activities applicable to the practice should use it, or else justify why a different practice is appropriate.

In addition to the status of the endorsement, the process defines two kinds of applicability, they are:

- **Community Standard:** This standard is appropriate within a particular community, without any implication or recommendation for its use in other contexts.
- **Core Standard:** This standard is crosscutting and applicable to all of NASA’s Earth science data systems regardless of their specific communities of interest.

D. Strategic Support for Standards

The proposed process will provide a cadre of well-documented and mature standards that are applicable to the specific needs of given communities of interest. The process will give opportunity for affected communities to comment on and anticipate the standards. Still, because the needs of different communities differ, translation and reformatting services will remain essential. Planning for development of such services is an important aspect of any standards strategy.

To ensure success of standards, it will also be necessary for NASA to provide both inreach and outreach resources. Even well documented standards can be misinterpreted, so help desk, validation activities, and educational materials will be required.

IV. CANDIDATE STANDARD EXAMPLE

How does the Standards Process work in practice? In July 2004, a member of the OPeNDAP group (group that originated the DAP 2) submitted the Data Access Protocol (DAP) 2 to the SPG for review as a candidate standard. DAP 2 is widely used in the oceanographic and climate communities as a data transport mechanism.

Upon receiving the RFC, SPG members performed an Initial Review. We checked for completeness according to the requirements of our governing document, RFC-003 [Instruction to Authors]. We looked also for clarity and coherence in the submission and applicability to NASA Earth science data

systems. This initial review resulted in a request to the authors to revise the document to clarify applicability.

After the DAP 2 document was revised, it was again submitted to the SPG. During a new Initial Review, the revised RFC was judged to have relevance to data systems interoperability among the NASA community and the RFC documented the specification and the relevance of the proposal to NASA clearly. We identified the RFC as “ESE-RFC-004”. We convened a Technical Working Group (TWG) to conduct public reviews of the DAP 2. A TWG chair and TWG members were selected from the members of the SPG.

Because, we anticipate a great variety of RFC’s, the Standards Endorsement Process requires that each TWG develop tailored criteria. The TWG must determine, in the context of the RFC, what implementation means, and what kinds of information must be gathered to assess the success of implementation. In this case, the TWG crafted a set of questions to guide reviewers for the Implementation Review phase. Because this RFC was a specification, the questions were designed to determine the accuracy, clarity, and completeness of the implementation specification for the DAP 2. We sent the request for comment on implementation of the DAP 2 (ESE-RFC-004) via email to over 2000 members of the NASA Earth science community of data systems users and operators. This email list was developed by the conflation of several existing NASA email lists. We tried to reach the broadest possible cross-section of NASA stakeholders. The TWG also gathered from NASA management and from professional knowledge, a list identified as key NASA stakeholders representing activities most likely to be knowledgeable about or impacted by the RFC. The TWG contacted each key stakeholder personally via email and telephone and invited comment on DAP 2 Implementation.

The Request for Comment on Implementation yielded many responses. Some revealed minor errors in the specification, some reviewers pointed out future enhancements for the protocol, a few respondents did not find DAP appropriate for their use, but most of the responses provided positive comments. The TWG agreed that there were several minor errors in the document and recommended correction. However, these errors were judged by the TWG to not affect the DAP 2 protocol itself or its related implementation but rather affected the quality of the documentation. All such comments were referred to the RFC author. The RFC author agreed that the errors should be corrected and did revise the RFC. The TWG also shared the comments that suggested future enhancements with the RFC author. This feedback was greatly appreciated. While the future enhancements were beyond the scope of the current DAP 2 specification, those comments were forwarded to the OPeNDAP group as potential future work.

The revised RFC was issued a new version number. The TWG judged that these changes did not require a completely new review. Instead, we sent the revised document to just those reviewers who identified the initial errors. We asked them to affirm that their concerns were addressed. The reviewers found that all errors were corrected properly.

Our process requires that the SPG as a whole ratify the success or failure of community review. The TWG prepared a case and presented to the SPG. The TWG recommended, and the SPG agreed that the RFC should be promoted to Draft Standard status and that a review of the operational effectiveness of the specification could start.

The Operational Review proceeded much as the Implementation Review had. However, the information that the TWG needed to convince itself that DAP 2 was effective in an operational Earth science data systems context was somewhat different than what was identified for simple implementation. We again produced a set of questions to guide respondents. These questions were designed to discover issues of operational scaling, maintenance impact and so forth.

We again sent an email request to the same NASA Earth science community email list, including all those who had responded to our implementation appeal. But, the responses to this general appeal did not yield sufficient information for the TWG to be comfortable with judging operational impact. In the opinion of the TWG, the responses were too general. Most reviewers did not answer the specific questions but just sent general statements of endorsement. The TWG could not determine from these comments, the operational complexity of DAP or of the installations.

To develop more detail, the TWG revised the Operational Review guidance with additional questions designed to elicit metrics that might to determine the extent of use by users and how usage had changed over time. We also changed our technique for gathering responses. Each TWG member was assigned to personally contact individuals that had indicated in either the implementation or operation responses that they had experience with an operational instance of DAP. We contacted these individuals directly by email or by telephone and personally asked for this more detailed information. This second Operational Review yielded a set of much more comprehensive and detailed comments, mostly very positive.

Based on this Implementation Review the TWG again made a case to the SPG. The SPG recommended that the DAP 2 be endorsed as a NASA Earth science Community Standard. This recommendation was forwarded to NASA management.

V. TECHNICAL NOTE EXAMPLE

When the SPG first started its work in January 2004, the concept of a Technical Note was somewhat nebulous and thought to be a catch all for anything that was not a candidate standard. After some initial experience, we think this category of RFCs will be key to generating community driven standards.

To generate community driven standards, an outreach to communities of practice must occur to inform them about the Standards Process for science data systems and its benefit to science research and applications in their own community.

We first discussed using the process with NASA stakeholders in the global precipitation science community. These members expressed a need to define the science content of level 2 global precipitation products. There were several international as well as NASA meetings where they could

discuss what this type of product should contain with different groups of scientists. However, there was no organization or group that would support their efforts to define the product and get agreement from their community. We suggested that they write a Technical Note that defines the science content of level 2 global precipitation products and submit it as an RFC to the SPG. The SPG would work with the community to define the list of reviewers for this Technical Note and send out review notice and collect the reviews. The Technical Note would be publicly accessible on the SPG website. The community leaders for this effort could then concentrate on identifying key stakeholders for the review and to work on getting agreement on the Technical Note contents by revising the Technical Note in response to the reviews. After the Technical Note gets wide spread agreement, the community leaders could resubmit the Technical Note as a candidate standard RFC. The process support that the SPG provides may be just enough to encourage this kind of community driven standards to be produced.

We had a similar discussion with some land scientists who identified a need to standardize the science content of a surface reflectance product. Again, while there were multiple international and agency meetings where they could discuss this, again there was no organization that would support their efforts to standardize on this product. And without the process support, the scientists felt that the entire standardization process was beyond what they could do personally. However, when we discussed the use of a Technical Note and its review process as a pre-cursor to submission as a Candidate Standard, the scientists decided to start this process.

VI. IMPACT

The SPG and its process are designed to provide NASA with the best possible practices for NASA data systems and for serving NASA data to our stakeholders. These practices will also enhance external interoperability. For example, the Global Earth Observing System of Systems (GEOSS) envisions that systems from many nations will be linked through use of open standards at their interfaces. The SPG Standards Process can inform NASA's contribution of technologies and practices to these collaborations. The themes of community consultation, evidence gathering and open methodology are key to the development of working standards and are at the heart of the SPG process. The SPG and the standards endorsement process as a way of vetting proposed practices provides a confident basis for NASA's standards choices, ensuring that NASA's data systems investments are effectively managed.

REFERENCES

- [1] NASA Earth System Science and Applications Advisory Committee, "NewDISS: A 6- to 10-year Approach to Data Systems and Services for NASA's Earth Science Enterprise," February 2001, NASA document, unpublished, available from <http://eos.nasa.gov/seeds/>.
- [2] SEEDS Formulation Team, "Strategic Evolution of Earth Science Enterprise Data Systems (SEEDS) Formulation Team Final Recommendations Report," July 2003, NASA document, unpublished, available from <http://eos.nasa.gov/seeds/>.
- [3] NASA Standards Process Group, NASA website, from <http://spg.gsfc.nasa.gov>